∽ CORRECTION DU TRAVAIL MAISON CHAP 7 - G ∾
ECHANTILLONNAGE ET STATISTIQUES

Exercice 1. Une entreprise, spécialisé dans la fabrication d'ampoules, cherche à améliorer la qualité de sa production et affirme qu'il n'y a pas plus de 6% d'ampoules défectueuses dans sa production.

Une association de consommateurs réalise un test sur un échantillon et obtient 71 ampoules défectueuses sur 1 000.

1. Dans le cas où il y aurait exactement 6% d'ampoules défectueuses, déterminer un intervalle de fluctuation asymptotique au seuil de 95% de la fréquence d'ampoules défectueuses sur un échantillon aléatoire de taille 1 000.

   $p = 0,06$ et l'association de consommateurs a produit un échantillon de taille $n = 1000$.
   Les conditions $n \geq 30$, $np = 1000 \times 0.06 \geq 5$ et $n(1-p) = 1000 \times 0.94 \geq 5$ sont vérifiées, il y a donc un sens à calculer l'intervalle de fluctuation au seuil 0.95

   $$p - \frac{1}{\sqrt{n}} = 0.06 - \frac{1}{\sqrt{1000}} \simeq 0.028 \qquad \text{et} \qquad p + \frac{1}{\sqrt{n}} = 0.06 + \frac{1}{\sqrt{1000}} \simeq 0.092$$

   L'intervalle de fluctuation au seuil 0.95 est $[0.028; 0.092]$

2. A-t-on des raisons de remettre en cause l'affirmation de l'entreprise ?

   On teste l'hypothèse il n'y a pas plus de 6% d'ampoules défectueuses dans la production de l'entreprise.
   La fréquence qu'a pu observé l'association de consommateur est $f = 0.071$. $f \in [0.028; 0.092]$, l'association de consommateur ne peut pas remettre en cause l'affirmation de cette entreprise.

Exercice 2. Afin d'estimer la proportion, pour l'année 2020, de personnes diagnostiquées diabétiques dans la population française âgée de 20 à 79 ans, on interroge au hasard 10 000 personnes. Dans l'échantillon étudié, 716 personnes ont été diagnostiquées diabétiques.

1. À l'aide d'un intervalle de confiance au niveau de confiance 95%, estimer la proportion de personnes diagnostiquées diabétiques dans la population française âgée de 20 à 79 ans. On ne connaît pas la proportion p de personnes diabétiques donc a produit un échantillon de taille $n = 10000$ afin de l'estimer à partir de la fréquence $f = \frac{716}{10000} = 0.0716$ que nous avons observé.
   Calculons les deux bornes de notre intervalle de confiance :

   $$f - \frac{1}{\sqrt{n}} = 0.0716 - \frac{1}{\sqrt{10000}} = 0.0616 \qquad \text{et} \qquad f + \frac{1}{\sqrt{10000}} = 0.0816$$

   Nous estimons notre proportion p de diabétiques en France entre 20 et 79 ans qui vérifie (à 95%) :

   $$0.0616 \leq p \leq 0.0816$$

2. Quel doit être le nombre minimal de personnes à interroger si l'on veut obtenir un intervalle de confiance d'amplitude inférieure ou égale à 0,01 ?

   Définition *: L'amplitude d'un intervalle est la différence entre la borne supérieure et la borne inférieur de l'intervalle. Par exemple, l'amplitude de l'intervalle $[a; b]$ vaut $b - a$.*

   On souhaite donc que la différence entre les deux bornes de notre intervalle de confiance soit inférieur ou égale à 0.01 autrement dit on cherche le plus petit entier naturel n tel que :

   $$f + \frac{1}{\sqrt{n}} - \left(f - \frac{1}{\sqrt{n}}\right) \leq 0.01 \Longleftrightarrow f + \frac{1}{\sqrt{n}} - f + \frac{1}{\sqrt{n}} \leq 0.01$$

   et donc

   $$\frac{2}{\sqrt{n}} \leq 0.01 \Longleftrightarrow 2 \leq 0.01 \times \sqrt{n} \Longleftrightarrow \frac{2}{0.01} \leq \sqrt{n} \Longleftrightarrow 200 \leq \sqrt{n}$$

   et au final en élevant au carré les deux nombres positifs on obtient :

   $$40000 \leq n$$

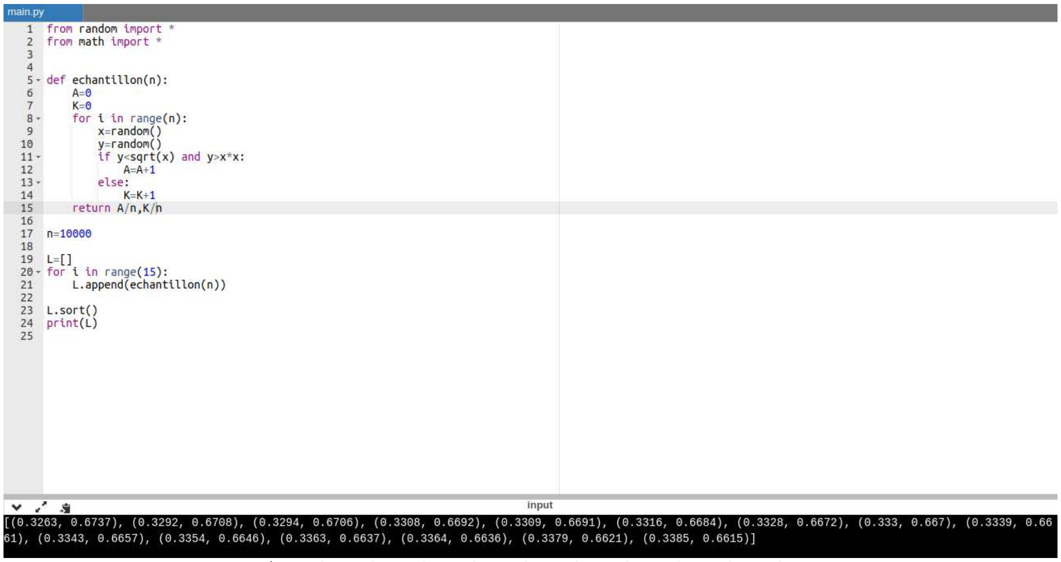
   Il faudra donc interrogé au moins 40000 personnes pour obtenir une estimation de p d'amplitude inférieure ou égale à 0.01.

Exercice 3.

1. Monsieur A gagne X euros, monsieur B gagne X^2 et monsieur C gagne X^3 euros par mois.
   (a) Calculer le salaire médian et le salaire moyen des trois personnes lorsque $X = 1$. Donner l'écart entre le salaire médian et le salaire moyen. Si $X = 1$ chacun gagne 1 euro par mois et le salaire médian et moyen vaut donc 1. L'écart entre le salaire médian et le salaire moyen est donc de 0.
   (b) Calculer le salaire médian et le salaire moyen des trois personnes lorsque $X = 10$. Donner l'écart entre le salaire médian et le salaire moyen.
   Le salaire moyen est :
   $$m = \frac{10 + 100 + 1000}{3} = 370$$
   Le salaire médian est 100, le salaire moyen est supérieur au salaire médian de 270 euros.
2. Le salaire net médian mesuré en équivalent temps plein par l'Insee se monte à 1.789 euros par mois selon les dernières statistiques disponibles .
   Le salaire médian net est inférieur de 20,1% au salaire moyen net des prélèvements à la source (CSG, CRDS notamment), qui se monte à 2.238 euros toujours selon l'Insee
   (a) Démontrer que, comme annoncé dans l'énoncé, le salaire médian net est inférieur de 20.1% au salaire moyen net.
   La différence entre les deux salaires vaut $2238 - 1789 = 449$ ce qui rapporté au salaire moyen donne un taux
   $$t = \frac{449}{2238} \simeq 20.1\%$$
   Autrement dit en diminuant le salaire moyen de 20,1% on obtient le salaire médian.
   En effet :
   $$2238 \times \left(1 - \frac{20.1}{100}\right) \simeq 1789$$
   (b) Comment expliquez une telle différence entre le salaire médian et le salaire moyen ?
   La moyenne est un indicateur sensible "aux valeurs extrêmes" d'une série statistique contrairement à la médiane. Observez une moyenne nettement plus élevé que la médiane traduit le fait que les inégalités de salaires en France sont très grande; les rares qui gagnent beaucoup d'argent en gagne tellement qu'il faut s'élever la moyenne très au dessus de la médiane.

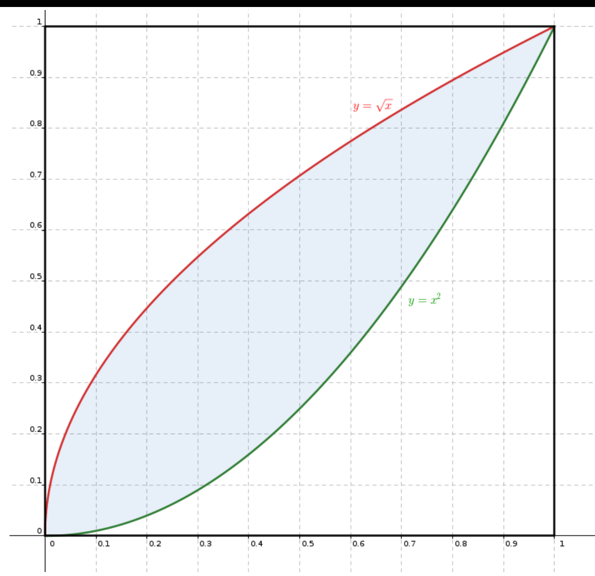
Exercice 4. **Simulation et méthode de Monté Carlo**

Compléter et reproduire sur python l'algorithme suivant afin de **proposer** une **estimation** de l'aire entre les courbes de la fonction $f(x) = \sqrt{x}$ et $g(x) = x^2$ sur l'intervalle 0; 1].




```python
from random import *
from math import *


def echantillon(n):
    A=0
    K=0
    for i in range(n):
        x=random()
        y=random()
        if y<sqrt(x) and y>x*x:
            A=A+1
        else:
            K=K+1
    return A/n,K/n

n=10000

L=[]
for i in range(15):
    L.append(echantillon(n))

L.sort()
print(L)
```

```
[(0.3263, 0.6737), (0.3292, 0.6708), (0.3294, 0.6706), (0.3308, 0.6692), (0.3309, 0.6691), (0.3316, 0.6684), (0.3328, 0.6672), (0.333, 0.667), (0.3339, 0.66
61), (0.3343, 0.6657), (0.3354, 0.6646), (0.3363, 0.6637), (0.3364, 0.6636), (0.3379, 0.6621), (0.3385, 0.6615)]
```



On conjecture que l'aire entre les deux courbes vaut $\frac{1}{3}$.